Sélection de modèle pour la construction de bandes de confiance sur la fonction moyenne

Emilie Devijver et Adeline Samson,
emilie.devijver@univ-grenoble-alpes.fr

Outline

- Model
- When the level L^* is known
- When the level L^* is unknown: approximation by a fixed L
- When the level L^* is unknown: selecting the best L

Model

For $i = 1, \ldots, N$ and $t \in [0, 1]$,

$$Y_i(t) = f(t) + \varepsilon_i(t). \quad (1)$$

We observe a discretization of this process: for $j = 1, \ldots, n$,

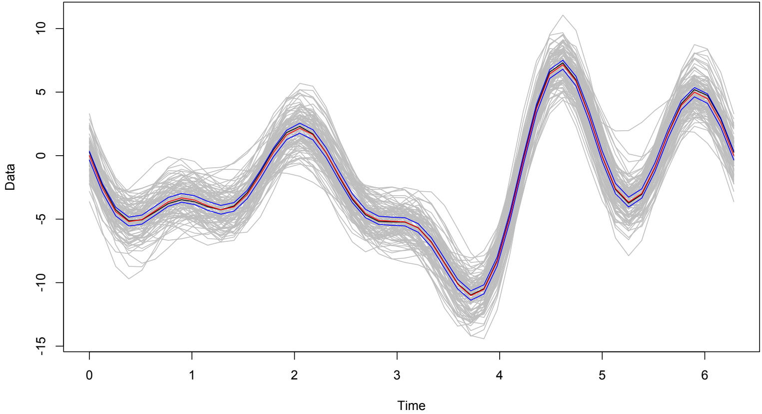
$$Y_{ij} = f(t_j) + \varepsilon_{ij}, \quad (2)$$

Goal: estimate f

- estimation: projection onto a functional basis
- control of this estimation: confidence band

Illustration

n = 300, N=50



Assumptions

- The function f belongs to a space of dimension L^*
$f \in Span((B_\ell^{L^*})_{1\leq \ell \leq L^*})$ and is denoted f^{L^*}:

$$f^{L^*}(t) = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t)$$

- $(B_\ell^{L^*})_{1\leq \ell}$ is an orthonormal basis
- The sequence ε_i is functional and belongs to $Span((B_\ell^{L^\varepsilon})_{1\leq \ell \leq L^\varepsilon})$. There exists a sequence of coefficients $c_{i\ell}$ such that

$$\varepsilon_{ij} = \sum_{\ell=1}^{L^\varepsilon} c_{i\ell} B_\ell^{L^\varepsilon}(t_j)$$

- The noise is Gaussian: for all $i = 1, \ldots, N$ and $\ell = 1, \ldots, L^\varepsilon$,

$$c_{i\ell} \sim_{iid} \mathcal{N}(0, \sigma^2)$$

When L^* is known

For a fixed $t \in [0, 1]$, the estimator of $f^{L^*}(t)$ is defined by:

$$\hat{\boldsymbol{\mu}}^{L^*} = (\mathrm{B}_{L^*}^T \mathrm{B}_{L^*})^{-1} \mathrm{B}_{L^*}^T \mathrm{Y}$$

$$\hat{f}^{L^*}(t) = \sum_{\ell=1}^{L^*} \hat{\mu}_\ell^{L^*} B_\ell^{L^*}(t)$$

with $\mathrm{B}_{L^*}^T = (B_\ell^{L^*}(t_j))_{1 \le \ell \le L^*, 1 \le j \le n}$.

When L^* is known

Proposition The distributions of the estimator $\hat{\boldsymbol{\mu}}^{L^*}$ and of the estimated function are

$$\begin{aligned} \hat{\boldsymbol{\mu}}^{L^*} &\sim \mathcal{N}\left(\boldsymbol{\mu}^{L^*}, \sigma^2 \Sigma_B^{L,L^\varepsilon}\right) \\ \hat{f}^{L^*}(t) - f^{L^*}(t) &\sim \mathcal{N}\left(0, \sigma^2 B(t) \Sigma_B^{L^*,L^\varepsilon} B(t)^T\right) \end{aligned}$$

where

$$\begin{aligned} \Sigma_B^{L^*,L^\varepsilon} &= (\mathrm{B}_{L^*}^T \mathrm{B}_{L^*})^{-1} \mathrm{B}_{L^*}^T \Sigma_{L^\varepsilon} \mathrm{B}_{L^*} (\mathrm{B}_{L^*}^T \mathrm{B}_{L^*})^{-1} \\ \Sigma_{L^\varepsilon} &= Diag(\Sigma_{L^\varepsilon}, \ldots, \Sigma_{L^\varepsilon}) \\ \Sigma_{L^\varepsilon} &= \mathrm{B}_{L^\varepsilon} \mathrm{B}_{L^\varepsilon}^T. \end{aligned}$$

When L^* is known

For a fixed confidence level α, we are looking for d^{L^*} such that

$$P\left(\forall t \in [0,1],\ \hat{f}^{L^*}(t) - d^{L^*}(t) \leq f^{L^*}(t) \leq \hat{f}^{L^*}(t) + d^{L^*}(t)\right) = 1-\alpha$$

Constant band

$$P\left(\max_{t\in[0,1]} |\hat{f}^{L^*}(t) - f^{L^*}(t)| \geq d^{L^*}\right) = \alpha$$

When L^* is known

Kac-Rice formulae

- If $X = \{X(t), t \in [0,1]\}$ is a centered Gaussian process with variance σ^2, $\mathcal{C}^1([0,1])$ almost surely
- Let $\tau(t)^2 = \frac{Var(X'(t))}{\sigma^2}$.

Then

$$\mathbb{P}(\exists t \in [0,1] : X(t) \geq d) \leq \Phi\left(\frac{-d}{\sigma}\right) + \frac{\|\tau\|_1}{2\pi} \exp\left\{-\frac{d^2}{2\sigma^2}\right\}.$$

Looking for d such that

$$\Phi\left(\frac{-d}{\sigma}\right) + \frac{\|\tau\|_1}{2\pi} \exp\left\{-\frac{d^2}{2\sigma^2}\right\} = \alpha.$$

Not constant band

Extension using Liebl and Reimherr[1] when the band is adaptive

[1] Fast and Fair Simultaneous Confidence Bands for Functional Parameters, arXiv 2022

When L^* is known

Experiments: confidence band for f^{L^*}

Setting

Fourier basis, of level $L^* = 11$,
$\mu_\ell \sim \mathcal{U}(\{-5, -4, -3, -2, 2, 3, 4, 5\})$ for all $\ell \in \{1, \dots, L^*\}$

Results

α / (n,N)	(50, 100)	(500, 1000)
0.05	0.947	0.948
0.1	0.914	0.905
0.2	0.842	0.812

Table: Repetitions over 1000 iterations for the constant band

Conclusion

Good coverage, asymptotics

When L^* is unknown: for a fixed L

Let L. We denote

$$\mu_\ell^{L,L^*} =< f^{L^*}, B_\ell^L > \text{ for } \ell \in \{1, \ldots, L\}$$

$$f^{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t) \mu_\ell^{L,L^*}$$

When L^* is unknown: for a fixed L

Let L. We denote

$$\mu_\ell^{L,L^*} = < f^{L^*}, B_\ell^L > \text{ for } \ell \in \{1, \ldots, L\}$$

$$f^{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t) \mu_\ell^{L,L^*}$$

But we are not able to observe the functions on infinite set of points:

$$\mathrm{m}^{L,L^*} = (\mathrm{B}_L^T \mathrm{B}_L)^{-1} \mathrm{B}_L^T \mathrm{B}_{L^*} \mu^{L^*}$$

$$\underline{f}^{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t) m_\ell^{L,L^*}$$

When L^* is unknown: for a fixed L

Let L. We denote

$$\mu_\ell^{L,L^*} =< f^{L^*}, B_\ell^L > \text{ for } \ell \in \{1, \dots, L\}$$

$$f^{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t)\mu_\ell^{L,L^*}$$

But we are not able to observe the functions on infinite set of points:

$$\mathrm{m}^{L,L^*} = (\mathrm{B}_L^T \mathrm{B}_L)^{-1} \mathrm{B}_L^T \mathrm{B}_{L^*} \mu^{L^*}$$

$$\underline{f}^{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t) m_\ell^{L,L^*}$$

When $n \to +\infty$,

$$\mathrm{m}^{L,L^*} \to \mu^{L,L^*} \text{ and } \underline{f}^{L,L^*}(t) \to f^{L,L^*}(t) \text{ for all } t$$

When L^* is unknown: for a fixed L

Let L.
For a fixed $t \in [0,1]$, the estimator of $\underline{f}^{L,L^*}(t)$ is defined by:

$$\hat{m}^{L,L^*} = (\mathrm{B}_L^T \mathrm{B}_L)^{-1} \mathrm{B}_L^T \mathrm{y}$$

$$\hat{\underline{f}}^{L,L^*}(t) = \sum_{\ell=1}^{L} \hat{m}_\ell^{L,L^*} B_\ell^L(t)$$

When L^* is unknown: for a fixed L

Let L.
For a fixed $t \in [0,1]$, the estimator of $\underline{f}^{L,L^*}(t)$ is defined by:

$$\hat{m}^{L,L^*} = (\mathrm{B}_L^T \mathrm{B}_L)^{-1} \mathrm{B}_L^T \mathrm{y}$$

$$\hat{\underline{f}}^{L,L^*}(t) = \sum_{\ell=1}^{L} \hat{m}_\ell^{L,L^*} B_\ell^L(t)$$

$$\hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) \sim \mathcal{N}\left(0, \sigma^2 B(t) \Sigma_B^{L,L^\varepsilon} B(t)^T\right)$$

$$\hat{\underline{f}}^{L,L^*}(t) - f^{L^*}(t) \sim \mathcal{N}\left(b^{L,L^*}(t), \sigma^2 B(t) \Sigma_B^{L,L^\varepsilon} B(t)^T\right)$$

where

$$b^{L,L^*}(t) = \underline{f}^{L,L^*}(t) - f^{L^*}(t).$$

When L^* is unknown: for a fixed L

Constant band

$$P\left(\max_{t\in[0,1]} |\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)| \geq d^{L,L^*}\right) = \alpha$$

Looking for d^{L,L^*} such that

$$\Phi\left(\frac{-d^{L,L^*}}{\sigma}\right) + \frac{\|\tau_L\|_1}{2\pi}\exp\left\{-\frac{(d^{L,L^*})^2}{2\sigma^2}\right\} = \alpha.$$

When L^* is unknown: for a fixed L

Experiments: Confidence band for $\underline{f}^{L,L^*}$

Setting

Fourier basis, of level $L^* = 11$,
$\mu_\ell \sim \mathcal{U}(\{-5,-4,-3,-2,2,3,4,5\})$ for all $\ell \in \{1,\ldots,L^*\}$
$N = 100, n = 50$

Results

α / L	3	5	7	9	11	13	15
0.05	0.997	0.983	0.974	0.964	0.947	0.949	0.955
0.1	0.991	0.970	0.950	0.934	0.914	0.920	0.901
0.2	0.970	0.942	0.904	0.871	0.842	0.806	0.832

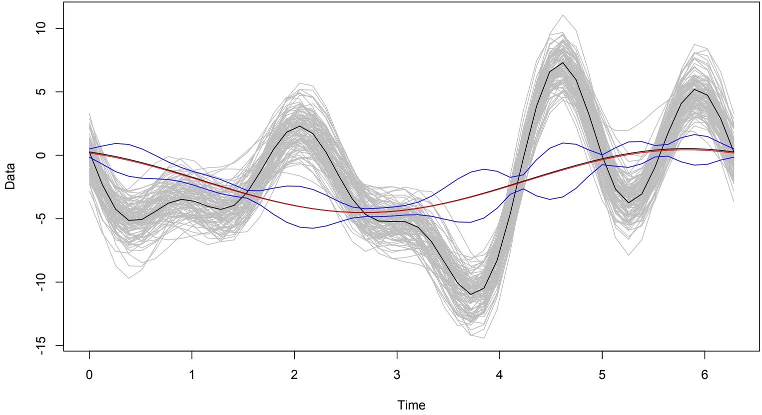
Table: Repetition over 1000 iterations, for $\underline{f}^{L,L^*}$.

Conclusion

Very conservative when $L \neq L^*$ for $\underline{f}^{L,L^*}$.

When L^* is unknown: for a fixed L

Experiments: Confidence band for $\underline{f}^{L,L^*}$



Conclusion

The estimation and the band are meaningless!

When L^* is unknown: for a fixed L

Confidence band for f^{L^*}

We want to control

$$\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) + \hat{f}^{L^*}(t) - \hat{f}^{L^*}(t) + \underline{\hat{f}}^{L,L^*}(t) - \underline{\hat{f}}^{L,L^*}(t) - f_{L^*}(t)$$

When L^* is unknown: for a fixed L

Confidence band for f^{L^*}

We want to control

$$\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) + \hat{f}^{L^*}(t) - \hat{f}^{L^*}(t) + \underline{\hat{f}}^{L,L^*}(t) - \underline{\hat{f}}^{L,L^*}(t) - f_{L^*}(t)$$

$$P\left(\max_{t\in[0,1]} |\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)| \geq d_1^{L,L^*}\right) = \alpha$$

When L^* is unknown: for a fixed L

Confidence band for f^{L^*}

We want to control

$$\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) + \hat{f}^{L^*}(t) - \hat{f}^{L^*}(t) + \underline{\hat{f}}^{L,L^*}(t) - \underline{\hat{f}}^{L,L^*}(t) - f_{L^*}(t)$$

$$P\left(\max_{t\in[0,1]} |\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)| \geq d_1^{L,L^*}\right) = \alpha$$

$$P\left(\max_{t\in[0,1]} |\underline{\hat{f}}^{L,L^*}(t) - \hat{f}^{L^*}(t) - \underline{f}^{L,L^*}(t) + f^{L^*}(t)| \geq d_2^{L,L^*}\right) = \alpha$$

When L^* is unknown: for a fixed L

Confidence band for f^{L^*}

We want to control

$$\hat{\underline{f}}^{L,L^*}(t) - f^{L^*}(t) = \hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) + \hat{f}^{L^*}(t) - \hat{f}^{L^*}(t) + \hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L,L^*}(t) - f_{L^*}(t)$$

$$P\left(\max_{t\in[0,1]} |\hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)| \geq d_1^{L,L^*}\right) = \alpha$$

$$P\left(\max_{t\in[0,1]} |\hat{\underline{f}}^{L,L^*}(t) - \hat{f}^{L^*}(t) - \underline{f}^{L,L^*}(t) + f^{L^*}(t)| \geq d_2^{L,L^*}\right) = \alpha$$

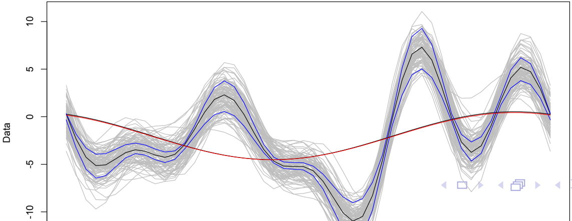
$$\hat{b}^{L,L^*}(t) = \hat{\underline{f}}^{L,L^*}(t) - \hat{f}^{L^*}(t)$$

When L^* is unknown: for a fixed L

Confidence band for f^{L^*}

$$P\left(\forall t \in [0,1], \underline{\hat{f}}^{L,L^*}(t) - \hat{b}^{L,L^*}(t) - f^{L^*}(t) \notin [-d_1^{L,L^*} + d_2^{L,L^*}, d_1^{L,L^*} - d_2^{L,L^*}]\right) = 1 - (1-\alpha)^2$$

- Terms in L^* may be approximated by $L_{\max}$ large
- Band centered in $\hat{f}^{L^*}$
- Remark: comparison with a confidence band at level $L_{\max}$ leads to the same width, but many more parameters!



When L^* is unknown: selecting the best L

Criterion: looking for the smaller band

$$\underset{L}{\operatorname{argmin}}\{d_1^{L,L^*} - d_2^{L,L^*}\}$$

α / L	3	5	7	9	11	13	15
0.05	2.66	2.23	1.67	0.89	0.51	0.52	0.51
0.1	2.47	2.07	1.54	0.83	0.48	0.48	0.48
0.2	2.27	1.90	1.41	0.75	0.44	0.44	0.44

Table: Repetition over 1000 iterations, for $\underline{f}^{L,L^*}$ debiased.

Model selection: penalize by the dimension to promote models of smaller dimension

When L^* is unknown: selecting the best L

Classical tools: estimation and error approximation in ℓ_2 norm[2]
In our case: norm max

Ongoing work: adapting Lacour et al.[3] to our case, with estimation of the bias (and variance associated to this estimation)

[2] E. Brunel's HdR, 2013

[3] *Estimator selection: a new method with applications to kernel density estimation*, Sankhya 1, 2017

Conclusion and perspective

- Functional model on an orthonomal basis
- If the dimension is known: every is easy
- If we use a fixed dimension: theoretically, we can control everything; but in practice, can be meaningless
- Selection of the dimension to construct the smallest meaningful confidence band

Conclusion and perspective

- Functional model on an orthonomal basis
- If the dimension is known: every is easy
- If we use a fixed dimension: theoretically, we can control everything; but in practice, can be meaningless
- Selection of the dimension to construct the smallest meaningful confidence band

- Work on the theoretical result for the model selection
- Try on benchmark and real dataset

References

- Liebl and Reimherr, *Fast and Fair Simultaneous Confidence Bands for Functional Parameters*, arXiv, 2022
- Sun and Loader, *Simultaneous Confidence Bands for Linear Regression and Smoothing*, Annals of Statistics, 1994
- Goldenshluger and Lepski, *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality*, Annals of Statistics, 2011
- Brunel, *Contributions à l'estimation non-paramétrique adaptative dans les modèles de durées*. Mémoire de synthèse, HDR Université de Montpellier, 2013
- Lacour, Massart and Rivoirard, *Estimator selection: a new method with applications to kernel density estimation*, Sankhya 1, 2017